IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 26, 2018, accepted November 19, 2018, date of publication November 27, 2018,
date of current version December 27, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2883583

Multiple Classifier Fusion and Optimization for Automatic Focal Cortical Dysplasia Detection on Magnetic Resonance Images

XIAOXIA QU[1,2], JIAN YANG[1], LJILJANA PLATIŠ[3], ASLI KUMCU[3], DANNI AI[1], BART GOOSSENS[3], TINGZHU BAI[1], YONGTIAN WANG[1], JING SUI[4], KAREL DEBLAERE[5], AND WILFRIED PHILIPS[3], (Senior Member, IEEE)

[1]Beijing Engineering Research Center of Mixed Reality and Advanced Display, School of Optics and Photonics, Beijing Institute of Technology, Beijing 100081, China
[2]Radiology Department, Beijing Tongren Hospital, Capital Medical University, Beijing 100730, China
[3]Department of Telecommunications and Information Processing (imec-IPI-TELIN), Ghent University, 9000 Ghent, Belgium
[4]National Laboratory of Pattern Recognition, Brainnetome Center, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[5]Department of Radiology, Ghent University Hospital, 9000 Ghent, Belgium

Corresponding author: Jian Yang (jyang@bit.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFC0107800, in part by the National Science Foundation Program of China under Grants 61672099, 81627803, 61501030, and 61527827, in part by the China Scholarship Council under Grant 201206030018, and in part by CSC from Ghent University under Grant 01SC0213.

ABSTRACT In magnetic resonance (MR) images, detection of focal cortical dysplasia (FCD) lesion as a main pathological cue of epilepsy is challenging because of the variability in the presentation of FCD lesions. Existing algorithms appear to have sufficient sensitivity in detecting lesions but also generate large numbers of false-positive (FP) results. In this paper, we propose a multiple classifier fusion and optimization schemes to automatically detect FCD lesions in MR images with reduced FPs through constructing an objective function based on the F-score. Thus, the proposed scheme obtains an improved tradeoff between minimizing FPs and maximizing true positives. The optimization is achieved by incorporating the genetic algorithm into the work scheme. Hence, the contribution of weighting coefficients to different classifications can be effectively determined. The resultant optimized weightings are applied to fuse the classification results. A set of six typical FCD features and six corresponding Z-score maps are evaluated through the mean F-score from multiple classifiers for each feature. From the experimental results, the proposed scheme can automatically detect FCD lesions in 9 out of 10 patients while correctly classifying 31 healthy controls. The proposed scheme acquires a lower FP rate and a higher F-score in comparison with two state-of-the-art methods.

INDEX TERMS Focal cortical dysplasia, magnetic resonance image, brain lesion detection, optimal weighted multiple classifiers, genetic algorithm.

I. INTRODUCTION

Epilepsy is a cerebral disease that triggers unexpected epileptic seizures, affects more than 65 million people worldwide [1], and carries the risk of death. Neuropathological research has identified focal cortical dysplasia (FCD), which is a type of cerebral malformation that occurs during neocortical growth, as being closely associated with the symptoms of approximately 30% of epileptic patients. For these patients, resective surgery is the only curative option and has been known to have a high success rate [1], [2]. Magnetic resonance image (MRI) is always used in the pre-surgical process for patients with epilepsy. The identification of a lesion in the MRI has been associated with a favorable seizure outcome after surgery [3].

FCD lesions are categorized into three types: Type I, Type II, and Type III [4]. Type I FCD lesions are isolated lesions, which comprise either the radial (FCD Type Ia) or tangential (FCD Type Ib) dyslamination of the neocortex, while the combination of both variants is classified as FCD Type Ic [5]. In FCD type I, gray matter signal abnormalities (in FLAIR and T2 weighted) are less common than in FCD type II, while the white matter signal abnormalities (in FLAIR, T2 weighted, and T1 weighted) occur equally in both FCD type I and FCD type II [6]. Type II FCD lesions are

typically extra-temporal and characterized both by architectural abnormalities and dysmorphic neurons. FCD Type IIa is characterized by dysmorphic neurons without balloon cells, while FCD Type IIb is characterized by dysmorphic neurons and balloon cells. FCD Type III refers to cortical lamination abnormalities associated with hippocampal sclerosis, vascular malformations, tumors and other principal lesion [4], [7].

In MR images, the primary characteristics of FCD Type II are the blurring of the gray matter/white matter (GM/WM) junctions, cortical thickening, cortical hyper-intensity, and transmantle sign [8]. Additional features include the cortical curvature, atrophy/enlargement, sulcal depth, and textural changes [7]. To date, clinical FCD detection has been conducted by radiologists. However, it is difficult and challenging to detect lesions by visual analysis because FCD lesions can be subtle. Thus, searching for possible lesions is inefficient and also subject to the diagnostic experience of medical professionals [9]. FCD lesions may not be recognized because they cause extremely subtle changes in the cortex [10].

Consequently, automatic methods to assist in lesion detection may improve the efficiency and sensitivity of FCD diagnosis from MR images. In addition, detection of a FCD lesion prior to surgery can contribute to performing a complete resection. The main approaches are voxel-based morphometry (VBM) and machine learning classifiers using voxel-based and surface-based features.

Several VBM methods have been proposed to improve the detectability of FCD [9], [11], [12]. Kassubek *et al.* [12] developed a VBM technique based on the SPM99 software and produced an extension image that compared the gray matter density map with that of healthy controls to identify lesions. Kassubek's method provides a valuable additional screening tool to highlight the abnormal gray matter arising from lesional regions. Huppertz *et al.* [9] also adopted a VBM technique based on SPM99 to create junction images. Their method highlights brain areas with blurring at the GM/WM junctions. Bonilha *et al.* [11] performed VBM analysis to detect gray-matter concentration abnormalities and evaluate the surgical performance of dysplastic tissue resection. Their results confirmed that most patients (10 out of 11) with FCD lesions exhibited a statistically significant excess of gray-matter concentration [11]. The results obtained by these VBM methods have revealed that the image intensity has significant value in distinguishing FCD lesions from healthy regions. VBM methods are promising in terms of highlighting the abnormal gray matter and blurred GM/WM junctions. However, the results of VBM methods depend on the control population used in the analysis, because VBM compares the images of patients with those of healthy controls. Moreover, various studies have demonstrated that the gray matter probability maps used for VBM analysis are not affected by intrinsic neuropathological cortical changes, owing to the lack of correlation between the gray matter probability values and the cortical neuropathological measures in seemingly normal gray matter [7]. Additionally, simple voxel-based comparisons have the least robustness against noisy voxels.

Several advanced classification methods encompassing multiple distinguishable voxel-based features have been proposed to improve the performance of lesion detection. These methods include the naïve Bayesian (NB) classifier [3], neural network (NN) classifier [13], support vector machine (SVM) [14], and linear discriminant analysis (LDA) classifier [15]. Antel *et al.* [3] proposed a training-based two-stage Bayesian classifier (TSBC) for FCD detection. In the first stage, the classifier was trained with the computational feature maps of cortical thickness, gradient, and relative intensity, which have been proposed by Bernasconi *et al.* [8]. Additionally, the voxels were classified into lesional and non-lesional areas. Voxels that were classified as lesional in the first stage were reassessed during a second stage, wherein the texture-based features computed from the gray-level co-occurrence matrices were used to reclassify these voxels as either lesional or GM. The TSBC method achieved a sensitivity of 85% (17/20 lesions from 18 patients), while flagging small false positive (FP) regions in 5 out of 18 patients. Moreover, the specificity in healthy controls was 100%. Loyek *et al.* [14] used SVM to investigate the potential of statistical features, gray-level co-occurrence, and gray-level run-length textures in detecting FCD lesions. Their results revealed that those features and textures can provide promising results for identifying lesions, with a sensitivity between 0.846 and 0.982 and a specificity between 0.734 and 0.922 at the voxel level across five patients. The main weakness of these methods is either the poor specificity (large number of FPs) both in patients and healthy controls, or a high specificity at the expense of a lower sensitivity (fewer TPs detected). Many of these methods also require a two-stage process to eliminate the numerous identified FP voxels [14], [15].

Machine learning classifiers using surface-based features have also been shown to be useful in FCD detection. Besson *et al.* [10] proposed an NN classifier for FCD detection. The classifier is driven by composite surface features incorporating the cortical thickness, gradient, mean intensity of intra-cortical surfaces, sulcal depth, and curvature. Their method correctly identified 89% of the FCD lesions (17/19), but also flagged approximately 6 FP regions per patient on average and 2.5 FP clusters for half of the healthy controls on average. Thesen *et al.* [16] carried out a detailed evaluation for clusters of abnormal cortical thickness, gray-white matter contrast, local gyrification, sulcal depth, Jacobian distance, and curvature by varying the threshold and smoothing parameters. Their method successfully distinguished patients with FCD, heterotopia, polymicrogyria, and encephalomalacia from the controls for 94% of the time. Hong *et al.* [15] developed a method based on the surface features (cortical thickness, sulcal depth, curvature, relative intensity, and gradient) combined with a linear discriminant analysis (SLDA) classifier to automate the detection of lesions in MR image-negative but histologically-confirmed FCD.

After a two-stage analysis consisting of vertex-wise classification followed by cluster-wise classification, the SLDA method achieved sensitivity of 74% (14/19) and specificity of 100%. Adler *et al.* [17] proposed novel surface features called ''doughnut'' maps to identify FCD in a pediatric cohort. Their method achieved a higher sensitivity (73%) when the novel features were included, and a lower sensitivity (59%) when only the established features such as cortical thickness, were considered.

Although previous studies [15] have successfully detected FCD lesions, false-positive results have often been generated in image classification. Typically, a size threshold is used to remove the false-positive results. However, this approach has a high risk of erroneously labeling small subtle lesions as negative. Lesion detection provides candidate regions and alerts the clinician to focus on those suspicious areas. Excessive false positive results will burden the clinician by increasing the interpretation time. Moreover, users tend to ignore a large portion of candidate results when these systems over abundantly generate false positive results, which leads to the system being underused [18], [19]. Thus, the objective of this study was to minimize the number of false positives.

Multi-classifier systems (MCS) can potentially increase the classification specificity, and have been widely applied in many research fields [20], [21]. MCS was initially proposed by Chow [20] to optimize the joint decision of binary classifiers using appropriately defined weights. Wozniak *et al.* [21] identified several advantages of MCS, such as robustness to small datasets (as is often the case in FCD) and the ability to outperform the best individual classifier.

Inspired by the advantages of MCS, we propose a multiple classifier fusion and optimization (MCFO) scheme for FCD lesion detection. A genetic algorithm [22] was used to optimize the weights of four individual classifiers and the decision threshold by using an objective function constructed to simultaneously maximize the true positives (TP) and the minimize FPs via the F-score statistic. The motivation for this study is discussed later in the text. Connected region analysis was used to generate candidate lesions from the voxel classification results. Six features were input into the classifiers. Typically, classification schemes focus on optimizing the accuracy, and considers both the number of TPs and true negatives (TN). However, the number of TN voxels is orders of magnitude greater than the number of TP voxels. Thereby, the efficiency of an objective function, which attempts to optimize for greater accuracy, is reduced.

In this application, the F-score is a more suitable objective function because it does not consider the TNs. The main advantages of the proposed MCFO scheme are as follows: (1) MCFO is a complete pipeline that takes the segmented T1-weighted MR images as input and outputs the candidate FCD lesion regions; (2) the number and type of individual classifiers is customizable, and the framework optimizes the contribution of each individual classifier; (3) the occurrence of false positives is minimized by using the F-score in the optimization of the objective function.

TABLE 1. Clinical details of the patients with Type II FCD lesions.

#	Sex/Age	Location (Hemisphere)	Volume (mm^3)
1	F/28	Parietal Lobe (R)	3700
2	M/49	Precentral Gyrus (R)	1230
3	M/11	Frontal Lobe (L)	1750
4	M/42	Frontal Lobe near Cingulate Gyrus (L)	500
5	F/29	Precentral Gyrus (R)	4500
6	F/46	Frontal Lobe (L)	1660
7	F/41	Frontal Lobe near Cerebral falx (L)	940
8	F/32	Frontal Lobe near Precentral Gyrus (L)	1640
9	M/40	Parietal Lobe (L)	2190
10	M/43	Frontal Lobe (R)	23670

II. MATERIALS AND METHODS

A. IMAGES AND GROUND TRUTH

We investigated the T1-weighted MR images of 10 patients with FCD lesions (five females and five males, with a mean $\pm$ standard deviation (SD) of 36 $\pm$ 11 years of age) and 31 healthy controls (19 females and 12 males, with mean $\pm$ SD: 32 $\pm$ 9 years of age) [23]–[25]. The clinical details of the patients with FCD lesions considered in this study are presented in Table 1.

Moreover, each subject, for both the patients and healthy controls, had one T1-weighted MR image. The images were acquired at the Ghent University Hospital using a Siemens 3T MR scanner. The use of this data for research purposes was approved by the Ghent University Hospital ethics committee. Sagittal slices were acquired with a repetition time of 2530 ms, echo time of 2.58 ms, and flip angle of 7°. The slice thickness was 0.9 mm, and the resolution of each slice was 0.8594 mm $\times$ 0.8594 mm. After three-dimensional (3D) image reconstruction, each scan comprised 256 $\times$ 256 $\times$ 176 voxels.

We selected patients who were pre-surgically diagnosed in the clinic using T1-weighted and FLAIR images after correlation with clinical symptoms. The lesions were histopathologically confirmed after surgery. The lesional type of all patients was FCD Type II. All of the images used in this study were obtained before surgery. The FCD lesions were delineated by a neuroradiologist (KD) using the active contour segmentation of the ITK-SNAP tool [26] on the axial views of the T1-weighted with the auxiliary information of the FLAIR MR images. The segmentations were manually adjusted as needed. Each delineation lasted between 10 and 30 minutes, depending on the size, extent, and complexity level of the lesion. The delineations on the T1-weighted images were used as the ground truth to train and test the FCD detection algorithms.

B. PREPROCESSING

Preprocessing consists of brain extraction, intensity non-uniformity correction, registration, and tissue segmentation. Apart from the brain, brain extraction removes the skull and other structures unrelated to the FCD lesions (e.g., the eye and neck regions). The tool created by Smith [27]

was used to extract the brain region. The non-uniformity of image intensity is a consequence of image acquisition and can cause segmentation algorithms to fail or underperform. Non-uniformity correction was conducted with the modified expectation–maximization algorithm [28].

Spatial varieties of brain images between different subjects can result in the irregular alignments of lesional regions and must be preliminarily regulated before classification. Hence, the images of all subjects were refined first by rigid and then by experienced affine registrations [29], [30]. The T1 MR reference image selected for registration is a widely applied standard image obtained by the Montreal Neurological Institute and is called the MNI152 brain[1] image. Registration to a standard space is necessary for the segmentation and computation of the Z-score maps.

In the registered images, the brain tissue is segmented into gray matter (GM), white matter (WM), and cerebral spinal fluid (CSF) on the basis of intensity information. Considering the partial volume effect, the brain image is partitioned into partial volume maps of GM, WM, and CSF by using the hidden Markov random field (MRF) model and the expectation–maximization algorithm [28], which is embedded in the FSL5.0 software and named FMRIB's Automated Segmentation Tool (FAST).[2] The value in each tissue map indicates the proportion of a given voxel belonging to that tissue class. The MRF beta values for the main segmentation phase is were set to 0.1. Four iterations were carried out for the removal of bias field. The FWHM for bias field smoothing in FAST is 20 mm.

C. FCD FEATURE MAP COMPUTATION

The preprocessed images were utilized to compute the six features and discriminate between the lesional and non-lesional regions. These feature images are expressed in the form of a feature vector defined as $\boldsymbol{F} = \{F_1, F_2, F_3, F_4, F_5, F_6\}$, where F_1 is the cortical thickness map, F_2 is the gradient map, F_3 is the relative intensity map, F_4 is the GM/WM boundary width map, F_5 is the curvature, and F_6 is the sulcal depth.

The cortical thickness map (F_1) is a revealing FCD feature depicting the increase of the cortical thickness. The method proposed by Qu *et al.* [25] was used to compute the cortical thickness map. For a given voxel in the GM located at $\boldsymbol{v}$, the iterative local search in the neighborhood method identifies the closest voxels located in the WM and the CSF, which are denoted by $\boldsymbol{w}(\boldsymbol{v})$ and $\boldsymbol{c}(\boldsymbol{v})$, respectively. Then the cortical thickness at the given voxel $\boldsymbol{v}$ is computed as $F_1(\boldsymbol{v}) = d(\boldsymbol{v}, \boldsymbol{w}(\boldsymbol{v})) + d(\boldsymbol{v}, \boldsymbol{c}(\boldsymbol{v}))$, where d is the Euclidean distance between the two voxel positions. An increase in the cortex thickness indicates the possible lesional regions.

The gradient map F_2 describes the blurring of the GM/WM junction. A 3D Gaussian kernel (FWHM=3 mm) was used to calculate the gradient magnitude over the entire image [3]. For the GM/WM junction, the lesional regions tended to contain more blurring in comparison with the non-lesional regions.

The relative intensity map F_3 proposed in [8] is analogous to the hyper-intensities within the lesional regions. This map is calculated as $F_3(\boldsymbol{v}) = 1 - \left|g(\boldsymbol{v}) - B_g\right| / B_g$, where $g(\boldsymbol{v})$ is the intensity at a given location $\boldsymbol{v}$, and B_g is the mean intensity value for all brains at the boundary between the GM and the WM regions.

The GM/WM boundary width map F_4 describes the broadened GM/WM junction [25] and is an alternative approach toward quantifying the blurring arising from the FCD lesion at the GM/WM junction. In the partial volume maps of the GM and WM, which were obtained from the preprocessed images, all voxels with values greater than 0 and less than (but not equal to) 1 are defined as voxels belonging to the GM/WM boundary region. In the neighborhood method [25], an iterative local search is also applied to estimate the width at each voxel position within the GM/WM boundary region. The FCD lesions are characterized by values greater than those in non-lesional regions. Moreover, larger values suggest the expansion of the GM/WM junction and also reflect the blurring on the boundary, which is typically observed in lesional regions.

The curvature feature F_5 is approximated on the basis of the GM/CSF surface [11], [31]. The curvature describes the folding patterns of the cortex, and the FCD lesions may cause local curvature changes [10], [15]. When the FCD lesions are located at the bottom of a given sulcus, the lesional region is centered at the highest curvature of its sulcal pial surface and spread along its walls [10]. When the lesions are on the vertex of the cortex, the curvature values are typically smaller than those in the non-lesional cortical vertices [10]. To adopt the curvature values in the voxel-based analysis, the calculation was extended throughout the entire GM along the path from the GM/CSF to the GM/WM. This path was obtained by computing the potential map of the GM regions using the Laplace method (Jones et al., 2000).

Finally, the sulcal depth F_6 [10], [13], [31], [32] was the shortest distance measured between the WM surface and the boundary of the brain mask, which was derived by brain extraction. The lesional regions tended to have greater sulcal depth than the non-lesional regions, owing to the increase of the cortical thickness.

D. Z-SCORE FEATURE IMAGE COMPUTATION

The computation of the Z-score map from each feature map F yields $\boldsymbol{Z} = \{Z_1, Z_2, Z_3, Z_4, Z_5, Z_6\}$. Each element in $\boldsymbol{Z}$ is the Z-score map of the respective feature map in $\boldsymbol{F}$. The abnormal (lesional) regions are expected to have higher Z-scores (greater deviation from healthy images) in comparison with normal regions.

The Z-score map is created by first computing the mean and SD within the local sliding windows of the healthy controls. The mean model of the healthy controls for the i^{th}

[1] http://neuro.debian.net/pkgs/fsl-mni152-templates.html
[2] http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FAST

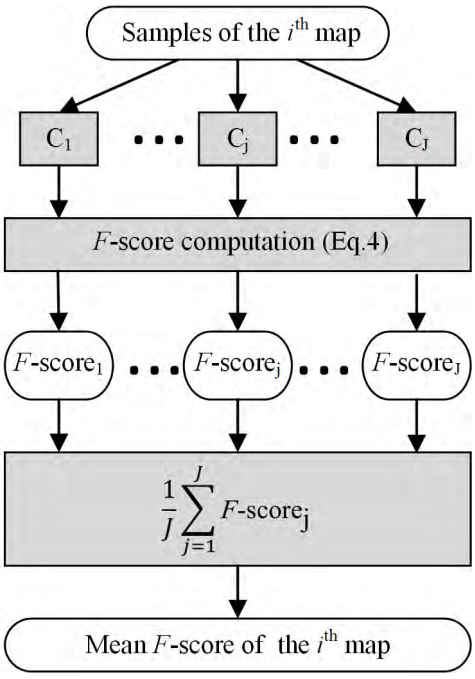

FIGURE 1. Computation of the mean F-score to identify the best performing maps. For each feature and Z-score map, an F-score is computed for each of the J classifiers. Then, the mean F-score per map is used to select the P best maps.

feature map F_i is $F_{\mu,i}(\boldsymbol{v}) = \frac{1}{K \cdot N}\sum_{k=1}^{K}\left(\sum_{n=1}^{N} F_i^{(k)}(\boldsymbol{v}_n)\right)$, where $F_{\mu,i}(\boldsymbol{v})$ is the mean model of the i^{th} feature map of all healthy controls at voxel $\boldsymbol{v}$. Moreover, $\boldsymbol{v}_n$ is the location of the n^{th} neighbor voxel within a 3D local window centered at $\boldsymbol{v}$. N is the total number of neighbor voxels. $F_i^{(k)}(\boldsymbol{v}_n)$ is the feature value of the k^{th} healthy control at voxel $\boldsymbol{v}_n$. K is the total number of healthy controls. The SD model of healthy controls for the i^{th} feature map is $F_{SD,i}(\boldsymbol{v}) = \frac{1}{\cdot N}\sum_{k=1}^{k}\left(\sqrt{\sum_{n=1}^{N}\left(F_i^{(k)}(\boldsymbol{v}_n) - F_{\mu,i}(\boldsymbol{v})\right)^2}\right)$. The Z-score map can be obtained by $Z_i(\boldsymbol{v}) = \left(F_i(\boldsymbol{v}) - F_{\mu,i}(\boldsymbol{v})\right) / \left(F_{SD,i}(\boldsymbol{v}) + \varepsilon\right)$, where $Z_i(\boldsymbol{v})$ is the value at voxel $\boldsymbol{v}$ on the i^{th} Z-score image and ε is set to 10^{-5}.

E. FCD FEATURE AND Z-SCORE MAPS SELECTION

The selection of the best performing maps (between the FCD feature maps $\boldsymbol{F}$ and the Z-score maps $\boldsymbol{Z}$) is guided by the procedure shown in Fig. 1. To train the classifiers for evaluating the maps, we combined the ground truth image and tissue segmented images for each subject, which resulted in an image with the labels of GM, WM, CSF, GM/WM, GM/CSF, and the FCD lesion. Using the labeled image, we extracted the sample values on each map at a spatial sampling interval of I_S. The I_S for each class (FCD, GM, and GM/WM junction) was adjustable to ensure that the total number of samples on each image maintained a constant value denoted by S_N.

For each feature, the samples were classified using multiple classifiers ($\boldsymbol{C} = \{C_1, \ldots, C_j, \ldots, C_J\}$), and the results were evaluated using the F-score, which is a measure of agreement between the classification result and the ground truth annotation of the lesion and will be introduced latter in Eq.4. For each $\boldsymbol{F}$ or $\boldsymbol{Z}$, the mean F-score over all classifiers is computed. Then, the feature and Z-score maps are ranked by the mean F-score. In the next step, during the training phase, the best-performing maps (those with highest F-scores) P are used to carry out F-score based optimization. During the testing phase, the best maps P, which are chosen during the training phase, are used to carry out F-score-based classification.

F. MULTIPLE CLASSIFIER FUSION AND OPTIMIZATION (MCFO)

1) OBJECTIVE FUNCTION

The procedure of multiple classifier fusion and optimization aims to maximize the F-score, which is the measure of the trade-off between *precision* and *recall* and can be expressed as $\mathrm{F}_{score} = 2 \times \text{precision} \times \text{recall} / (\text{precision} + \text{recall})$. In this context, precision (also called the positive predictive value) is defined as the number of true positive (TP) voxels divided by the total number of voxels identified as positive (TP voxels and false positive (FP) voxels) (precision $=$ #TP/ (#TP $+$ #FP)), where # indicates the number of voxels classified with that outcome. Recall (also called sensitivity or the true positive rate (TPR)) is the percentage of TP voxels with respect to all voxels that are positive according to the ground truth (total number of TP and false negative (FN) voxels) (recall $=$ #TP/ (#TP $+$ #FN)). Therefore, the F-score can be expressed as $\mathrm{F}_{score} = 2\,(\#\mathrm{TP}) \,/\, (2\,(\#\mathrm{TP}) + \#\mathrm{FN} + \#\mathrm{FP})$.

2) MULTIPLE CLASSIFIER FUSION

In the classification stage, each voxel is classified as lesion or normal using the j^{th} classifier applied to the P best performing maps. The j^{th} classifier gives the m^{th} voxel a label denoted by $l_j(\boldsymbol{v}_m)$. If the voxel is classified as lesional, $l_j(\boldsymbol{v}_m) = 1$; otherwise $l_j(\boldsymbol{v}_m) = 0$. Then, the classified labels are evaluated by the ground truth denoted by y $(\boldsymbol{v}_m)$, where $y(\boldsymbol{v}_m)$ is 1 if $\boldsymbol{v}_m \in FCD$; otherwise, $y(\boldsymbol{v}_m)$ is 0.

For the classifier fusion, a voxel is labeled as a TP only when the ground truth indicates that it is a lesional voxel, and the weighted sum of the classifier labels is above the decision threshold T. This means that TP $(\boldsymbol{v}_m) = 1$, if y $(\boldsymbol{v}_m) = 1$ and $\sum_{j=1}^{J} w_j l_j(\boldsymbol{v}_m) > T$; otherwise, TP $(\boldsymbol{v}_m) = 0$, where w_j is the weight of the j^{th} classifier, and J is the total number of classifiers. The weights are constrained to $\sum_{j=1}^{J} w_j = 1$; $0 \leq w_j \leq 1$. The labels for FN $(\boldsymbol{v}_m)$ and FP $(\boldsymbol{v}_m)$ are determined in a similar manner.

To fuse the classifiers, we apply the sign function. Alternatively TP $(\boldsymbol{v}_m)$ can be expressed as follows:

$$\mathrm{TP}(\boldsymbol{v}_m) = y(\boldsymbol{v}_m) \cdot \frac{\mathrm{h}(\boldsymbol{v}_m) + \mathrm{h}^2(\boldsymbol{v}_m)}{2} \tag{1}$$

where $\mathrm{h}(\boldsymbol{v}_m) = \mathrm{sgn}\left(\sum_{j=1}^{J} w_j l_j(\boldsymbol{v}_m) - T\right)$ is the sign function. $\mathrm{FN}(\boldsymbol{v}_m)$ and $\mathrm{FP}(\boldsymbol{v}_m)$ are deduced in the same manner; thereby, they yield the following equations:

$$\mathrm{FN}(\boldsymbol{v}_m) = y(\boldsymbol{v}_m)\left(1 - y(\boldsymbol{v}_m)\frac{\mathrm{h}(\boldsymbol{v}_m) + \mathrm{h}^2(\boldsymbol{v}_m)}{2}\right) \quad (2)$$

$$\mathrm{FP}(\boldsymbol{v}_m) = (1 - y(\boldsymbol{v}_m))\frac{\mathrm{h}(\boldsymbol{v}_m) + \mathrm{h}^2(\boldsymbol{v}_m)}{2}. \quad (3)$$

Accordingly, the F_{score} can be expanded as follows:

$$\mathrm{F}_{score} = \frac{2\sum_{m=1}^{M}\mathrm{TP}(\boldsymbol{v}_m)}{2\times\sum_{m=1}^{M}\mathrm{TP}(\boldsymbol{v}_m) + \sum_{m=1}^{M}\mathrm{FN}(\boldsymbol{v}_m) + \sum_{m=1}^{M}\mathrm{FP}(\boldsymbol{v}_m)} \quad (4)$$

where $\mathrm{TP}(\boldsymbol{v}_m), \mathrm{FN}(\boldsymbol{v}_m), \mathrm{FP}(\boldsymbol{v}_m)$ depend on $\mathrm{h}(\boldsymbol{v}_m)$, and $\mathrm{h}(\boldsymbol{v}_m)$ is computed from the classifier weightings (w_j), the classified labels by single classifiers ($l_j(\boldsymbol{v}_m)$), and the decision threshold (T). Therefore, the multiple classifiers are fused using Equation (4).

3) MULTIPLE CLASSIFIER OPTIMIZATION

To optimize the classifiers, the object function $\mathrm{obj}(w_j, T)$ is formulated as follows:

$$\mathrm{obj}(w_j, T) = \mathrm{F}_{score} - \left|\sum_{j=1}^{J} w_j - 1\right|. \quad (5)$$

where $0 \le w_j \le 1$. To determine the optimal weights of the multiple classifiers, the optimization is calculated as follows:

$$\mathrm{best}(w_j, T) = \mathrm{argmax}_{(w_j,T)}\left(\mathrm{obj}(w_j, T) \,|0 \le w_j \le 1\right) \quad (6)$$

The optimal weight w_j and threshold T maximizing the function $\mathrm{obj}(w_j, T)$ are determined by a genetic algorithm.

The genetic algorithm is implemented using the MATLAB toolbox[3] as reported by Houck *et al.* [22]. Considering that the input gene of the genetic algorithm is typically one variable in binary code, we encoded multiple decimal variables into one binary code sequentially. Each decimal variable was encoded into a binary code with a length of L_{BC}. The first decimal variable occupies the segment between 1 and L_{BC} of the binary code, while the second one is spread over the segment between $L_{\mathrm{BC}} + 1$ and $2 \times L_{\mathrm{BC}}$. Thus, the Q^{th} variable falls in the range of $(Q-1) \times L_{\mathrm{BC}} + 1$ to $Q \times L_{\mathrm{BC}}$. Additionally, L_{BC} is calculated in accordance with the range of the decimal variables $[a, b]$ and the dispersed precision denoted by ε_2 as follows: $L_{\mathrm{BC}} = \left\lfloor\left(\log_2\left(\frac{(b-a)}{\varepsilon_2} + 1\right)\right)\right\rfloor$, where $\lfloor\,\rfloor$ is the floor function truncating the supplied number down to the nearest integer. The outputs of the genetic algorithm are the optimal parameters for the voxel-based classification and are denoted as $\mathrm{v}_{best}(w_j, T)$ and $\mathrm{r}_{best}(w_j, T)$ in the regional classification.

4) CLASSIFICATION

The optimization of the voxel based classification is as follows: (1) voxels are classified as lesion or normal using the classifiers in $\boldsymbol{C}$; (2) the classified results are fused; (3) the weights are updated and the optimal voxel-based parameters $\mathrm{v}_{best}(w_j, T)$ are identified using the genetic algorithm; (4) the voxels are re-classified as lesion when $\sum_{j=1}^{J} w_j l_j(\boldsymbol{v}_m) > T$; otherwise they are classified as normal.

Then, the voxels labeled as positive are reclassified using the region based classification stage. First, if the voxels labeled as positive are connected, these connected voxels are considered as a single region. Then, the following region-based features are computed: mean, SD, skewness, and kurtosis, which have also been used in [15]. The skewness measures the asymmetry of the probability distribution of the regional feature values relative to the mean, whereas, kurtosis measures the features' probability distribution peak. Finally, the regions are classified as lesion or normal, and the region-based parameters are updated based on the genetic algorithm to determine the optimum parameters denoted as $\mathrm{r}_{best}(w_j, T)$.

G. EXPERIMENTAL DESIGN

The total number of neighboring voxels N for computing the Z-score images was set to 27. The best four maps were selected ($P = 4$) and the optimal P value was set experimentally. The number of samples in each image (S_N) was experimentally set to 500.

The individual classifiers $\boldsymbol{C} = \{C_1, \ldots, C_j, \ldots, C_J\}$ used in our experiment ($J = 4$) comprise the NB [33], [34], LDA, Mahalanobis discriminant analysis (MDA), and quadratic discriminant analysis (QDA) classifiers [35]. For the NB classifier, we use the normal (Gaussian) function as the distribution. In LDA, a multivariate normal density was applied to fit each class with a pooled covariance estimate. MDA uses Mahalanobis distances to compare the distances between the test samples and the training samples. QDA fits the multivariate normal densities with the covariance estimates stratified by the classes in the training data.

These classifiers calculate both the mean and the variance in each class and are stable in terms of convergence. By contrast, the classifiers based on the SVM or NN methods may fail to converge when the false positive rates are high. Moreover, in this particular application, the parametric estimation of the SVM and NN methods can be overly complicated. Therefore, the NB, LDA, MDA, and QDA classifiers were selected as a suitable set of classifiers for FCD detection. During the regional classification and evaluation, regions smaller than $R_{min} = 900$ voxels, which is equivalent to 112.5 mm^3 (0.125 mm^3/voxel), were considered as noisy fragments and were ignored.

The genetic algorithm parameters were set as follows: the population number was 50, the maximum generation was 100, the mutation method was a binary mutation with a probability of 0.05, the crossover method was a simple single-point crossover, and the crossover probability was 0.6. We determined the population and generation values empirically, as shown in Fig. 2. The population values between 10 and 200 and the generation values between 10 and 200 were tested. Equation 6, which includes the F-score,

[3]https://github.com/estsauver/GAOT

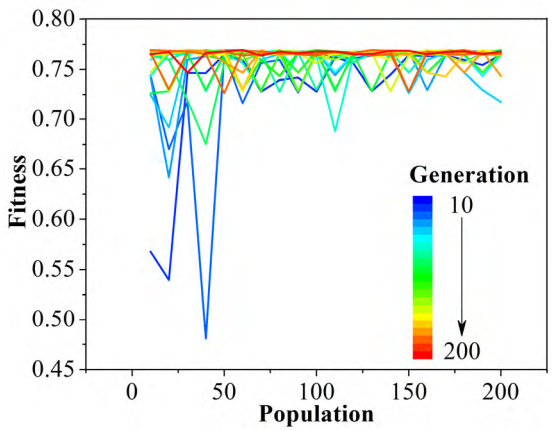

FIGURE 2. Performance of genetic algorithm parameters: population and generation. Each line corresponds to a trial that consists of one generation value. Equation 6, which includes F-score, is used as the measure of fitness.

was used as the measure of fitness. For the population, the fitness value stabilized between 0.70 and 0.75 starting at a population of 50. Although larger population values could be used, we chose the value of 50 because the performance achieved with this value was satisfactory. We selected 100 as the generation value, because it guarantees convergence without introducing an unnecessary computational cost. If the generation value is too small, the fitness function may be unable to successfully converge; if it is too large, the method will become time consuming. These experiments were conducted on a subset of the voxels in the image data. Therefore, the fitness values in Fig. 2 were larger than the F-score values, as will be presented in the Results section.

Considering the limited number of images in our dataset, the training and test images were evaluated with a leave-one-out cross validation strategy for all three algorithms. The classification results were evaluated at two levels. At the voxel level, the mean and SD of the following metrics were computed (equations given in Table 2): false positive rate (FPR), TPR (recall or sensitivity), precision (Pre), F-score, and accuracy (Acc). Because the review of a set of candidate FCD regions (instead of individual voxels) by medical experts is more efficient, we also evaluated the results at the regional level. A region is computed as a set of connected voxels and has a size greater than $R_{\min}$. For each patient, we reported that the lesion is detected when a region has at least one voxel overlapping with the ground truth region. The additional regions identified as positive do not contain a lesion and are false-positive regions.

H. IMPLEMENTATION OF PROPOSED METHOD, TSBC AND SLDA

The implementation of the proposed MCFO method and the features of the TSBC [3] and SLDA [15] can be found on Github website https://github.com/XiaoxiaQu/MRI-FCD-detection. For comparison with existing methods, the TSBC and SLDA were re-implemented according to the procedures and parameters reported in [3] and [14].

TABLE 2. The metrics applied for evaluation in this study.

Metric	Computational formula
Acc	(#TP + #TN)/(#TP + #TN + #FP + #FN)
TPR	#TP/(#TP + #FN)
FPR	#FP/(#FP + #TN)
Pre	#TP/(#TP + #FP)
F-score	2#TP/(2#TP + #FP + #FN)

To implement the TSBC method based on tissue segmented images, we computed the cortical thickness, gradient, relative intensity, and gray-level co-occurrence matrix (GLCM) textures in terms of volume. In the GLCM, the radius of the local window was set to 6, and the gray level was set to 32. The first stage Bayesian classifier used the cortical thickness, gradient, and relative intensity to classify the voxel into lesional and non-lesional. The voxels classified as lesional were then re-classified according to the GLCM textures using second stage Bayesian classifier.

For comparison with the SLDA methods in the same manner as TSBC, we also implemented the SLDA in the volume level, and not in the surface level. Although the SLDA method applies surface based features, some surface features were originally computed in terms of volume. Then, these features were mapped onto the cortical surface. To implement the SLDA method, we computed the features of the cortical thickness, sulcal depth, gradient, relative intensity in volume, and the surface level curvature. The curvature was then mapped into volume space for further classification. The LDA classification method was implemented in MATLAB R2014b using the classify function.

I. EVALUATION OF IMAGES BY EXPERTS

To compare the automated method with the classification carried out by experts, two radiologist from the Beijing Tongren Hospital of the Capital Medical University examined the T1 and FLAIR images. We included the data of 10 patients and 10 healthy controls. The two radiologists were blind to the ground truth. We considered that radiologists typically view T1, T2, and T2-FLAIR MRIs manually. However, this study mainly focused on T1 images. We designed two experiments. In the first experiment, the radiologists performed the diagnosis by only examining the T1 images, whereas in the second experiment, they performance the diagnosis by examining both T1 and FLAIR.

III. EXPERIMENTAL RESULTS

A. RESULTS OF FCD FEATURE AND Z-SCORE MAPS

All six feature maps and their corresponding Z-score maps are shown in Fig. 3 for a sample FCD patient. The mean feature maps of the healthy controls are also presented for comparison. Image A is an example T1-weighted MR image of a patient, wherein the FCD lesion is indicated with a red arrow. Image B is the T1-weighted MR image of a healthy

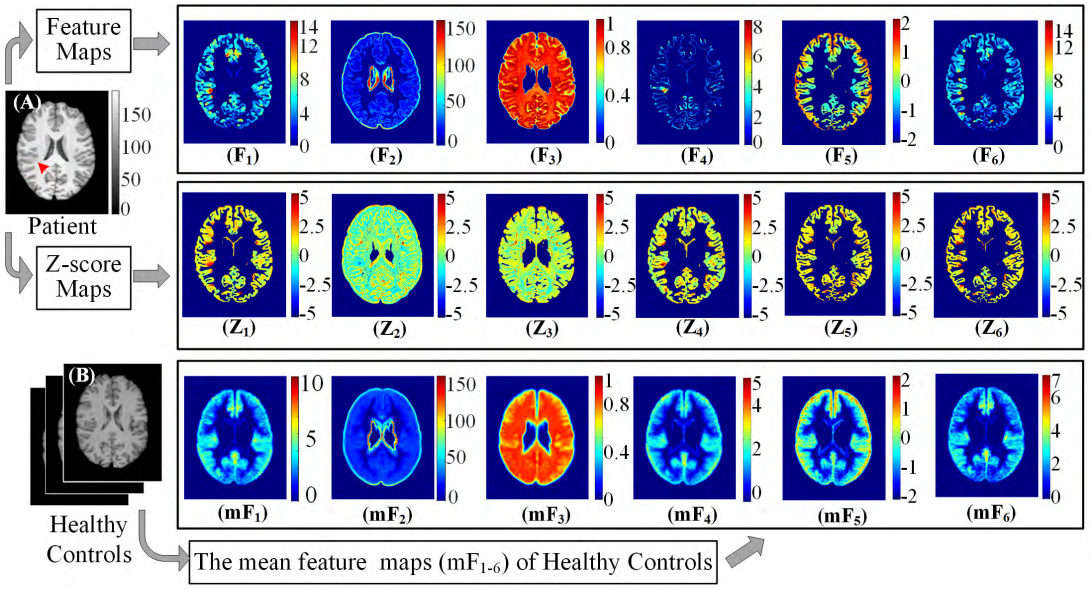

FIGURE 3. Experimental results of feature and Z-score maps for one patient and mean feature maps for healthy controls. (A) is the preprocessed T1-weighted MR image of the patient with lesional region, as indicated by a red arrow. Feature maps (F_{1-6}) are computed from (A) and are as follows: cortical thickness map (F_1, mm), gradient map (F_2, arbitrary unit), relative intensity map (F_3, arbitrary unit), gray/white matter boundary width map (F_4, mm), curvature (F_5, arbitrary unit) and sulcal depth (F_6, mm). Z-score maps (Z_{1-6}) of the patient correspond to the feature maps (F_{1-6}). An example slice from a healthy control is shown in (B), and corresponds to approximately the same slice as (A). The mean feature maps across all healthy controls are shown as in mF_{1-6}.

control, and the images on the right side of image B are the mean feature maps of the healthy controls. The six feature maps of (A) shown for the patient are as follows: cortical thickness (F_1), gradient (F_2), relative intensity (F_3), GM/WM boundary width (F_4), curvature (F_5), and sulcal depth (F_6). In comparison with the healthy regions, the lesional regions had higher cortical thickness values, higher GM/WM boundary width, higher relative intensity and sulcal depth, and lower gradient values. The curvature values in the lesional region depend on its location: that is, the lesional regions had higher curvature values in the vertex of the cortex and lower values in the sulci, in comparison with the corresponding non-lesional region in the healthy controls. In the mean feature maps, the value of each voxel was computed by averaging the voxel values at the same location of the healthy controls. Therefore, these maps appear to be much smoother than the maps of an individual.

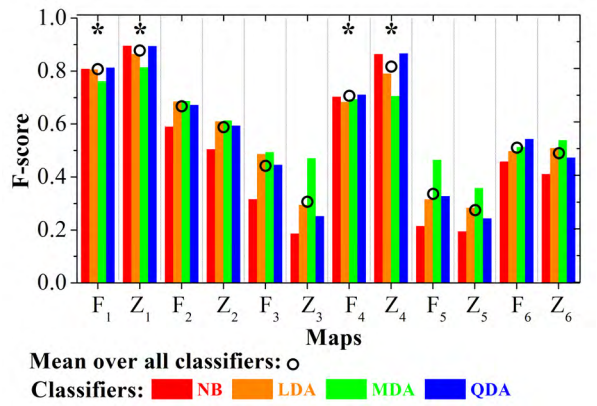

FIGURE 4. Evaluation results from the FCD feature and Z-score map selection (S4). The feature maps are cortical thickness map (F_1), gradient map (F_2), relative intensity map (F_3), gray/white matter boundary width map (F_4), curvatures (F_5) and sulcal depth (F_6). The Z_{1-6} are the corresponding Z-score image of F_{1-6}. An F-score per map was computed for each of the four classifiers, in which the black circles indicate the mean F-score over all classifiers. Maps F_1, F_4, Z_1 and Z_4 had the highest average performance (indicated by a $*$).

B. RESULTS OF FEATURE MAP EVALUATION

The four best performing maps in our experiment, as identified by the largest mean F-score (shown in Fig. 4), were the cortical thickness map (F_1), GM/WM boundary width map (F_4), Z-score map of cortical thickness (Z_1), and Z-score map of the GM/WM boundary width (Z_4). Subsequently, these four maps were used to train and test the MCFO classification scheme. The Z-score maps Z_1 and Z_4 performed better than the corresponding feature maps F_1 and F_4. However, for the other four FCD features, the Z-score maps performed worse than the corresponding feature maps. Overall, the relative intensity (F_3) and curvature (F_5) features and their corresponding Z-score maps had the lowest F-scores.

The $Z_{1,4}$ maps performed better than the $Z_{3,5}$ maps according to the evaluation results presented in Fig. 4, and the images are shown in Fig. 5. After comparing $F_{1,4,3,5}$ with the $mF_{1,4,3,5}$, the lesions in Z_1 and Z_4 were visible. However, the lesions in the Z_3 and Z_5 were not visible because they had

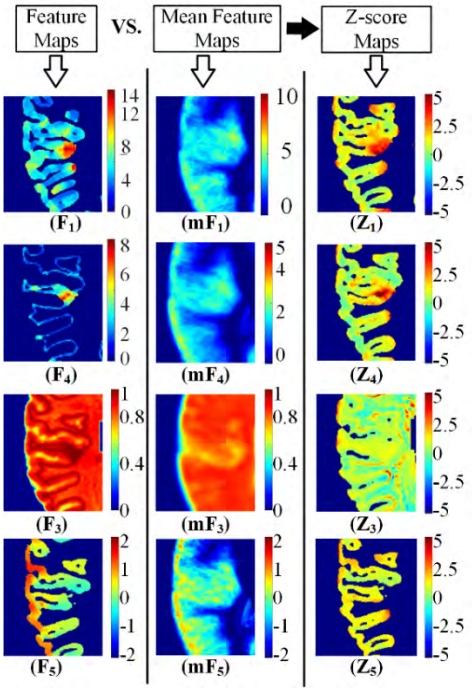

FIGURE 5. Illustration of enlarged lesional region on feature maps. $F_{1,4}$ and $Z_{1,4}$ outperform $F_{3,5}$ and $Z_{3,5}$. $F_{1,4,3,5}$ are compared with the healthy $mF_{1,4,3,5}$ maps, giving $Z_{1,4,3,5}$, respectively.

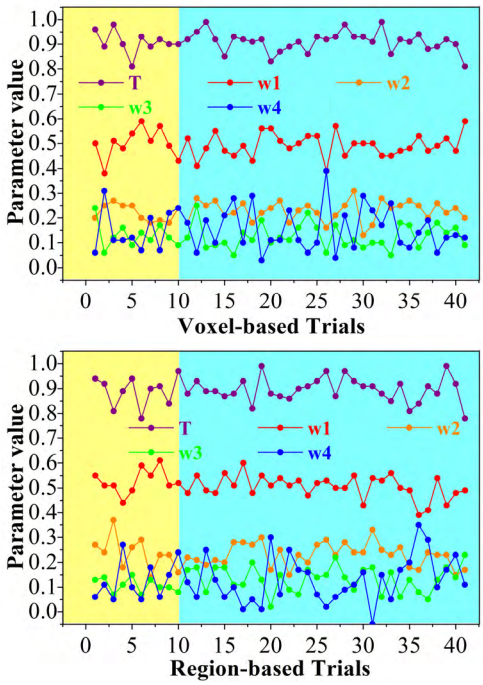

FIGURE 6. The optimal parameter values (threshold T and weights w_1-w_4) from F-score based optimization (step S5a) for each training trial. A total of 41 trials (the first 10 are FCD patients, followed by 31 healthy cases) were conducted with a leave-one-out validation scheme. The top plot shows the result of voxel-based classification and the bottom one of region-based classification.

similar values to the non-lesional regions. The two features that were identified as the most dominant in our experiment were the thickness of the gray matter (F_1), which is typically increased in the lesional areas, and the width between the gray and the white matter (F_4), which is also thicker in the lesional regions. The lesional regions in F_1 and F_4 had higher values than both of the non-lesional regions in F_1 and F_4, and mF_1 and mF_4. Therefore, the smooth mean value in the mF_1 and mF_4 did not decrease the contrast between the lesional and non-lesional regions.

The lesional region in F_3 was visually different from the non-lesional regions, as shown in Fig.5. However, the values in the lesional region are not unique. Along the entire GM/WM boundary, the values of 1 were observed to be the same as those in the lesional region. Thus, in the F_3 of the healthy control' images, all of the GM/WM boundary regions also had values close to 1. Thus, the mF_3 map also had high values in the regions wherein the GM/WM boundary accumulated from different healthy controls. When comparing F_3 to mF_3, the structural changes in the lesional region in F_3 could not be distinguished from those of the non-lesional region in Z_3.

In F_5, the differences between the lesional and non-lesional regions were not obvious. Although curvature has been applied to FCD detection in [10], [15], and [16], it has also been reported that the curvature performs worse than the other features [16]. This is probably caused by the fact that the cortex has a complex shape and does not have not locally constant curvature values. The convolved cortex of the different subjects is not necessarily the same in the healthy controls, but only approximately similar.

C. COMPARISONS OF CLASSIFICATION ALGORITHMS

The optimal parameters $v_{best}(w_j, T)$ and $r_{best}(w_j, T)$ of the MCFO scheme, which result from the training phase, are shown in Fig. 6. A total of 41 trials for each parameter are shown, because 41 images (10 patients and 31 healthy controls) were evaluated according to a leave-one-out validation scheme. Among the four classifiers, the weight of classifier 1 (w_1) was the largest, which suggests that NB outperformed LDA, QDA, and MDA. The w_j values were smaller than T, which suggests that a combination of multiple classifiers can outperform a single classifier.

D. COMPARISON OF FCD DETECTION ALGORITHMS

Three FCD detection algorithms were evaluated, namely, TSBC, SLDA, and the proposed MCFO scheme. From the voxel-based evaluation, which is summarized in Table 3, MCFO had a similar or better performance in comparison with SLDA, and both outperformed TSBC. The results of the regional analysis, which are presented in Table 4 and graphically in Fig. 7, revealed the same trends; that is MCFO and SLDA identified the lesion in 9 out of 10 patients, whereas, TSBC identified the lesion in 8 out of 10 patients and generated more FP regions in comparison with MCFO and SLDA. Figs. 8 and 9 present the detection results for the

TABLE 3. Evaluation of FCD detection algorithms using voxel-based analysis. Results are for patient images only. The values are mean (±SD).

	Accuracy	F score	Precision	TPR(recall)	FPR
TSBC	0.9445(±0.0275)	0.0708(±0.1131)	0.0427(±0.0751)	0.3933(±0.2754)	0.0536(±0.0270)
SLDA	0.9777(±0.0237)	0.2060(±0.1454)	0.1372(±0.1032)	**0.5374**(±0.2979)	0.0212(±0.0241)
MCFO	**0.9918**(±0.0051)	**0.3008**(±0.2269)	**0.2391**(±0.2016)	0.4500(±0.2822)	**0.0064**(±0.0040)

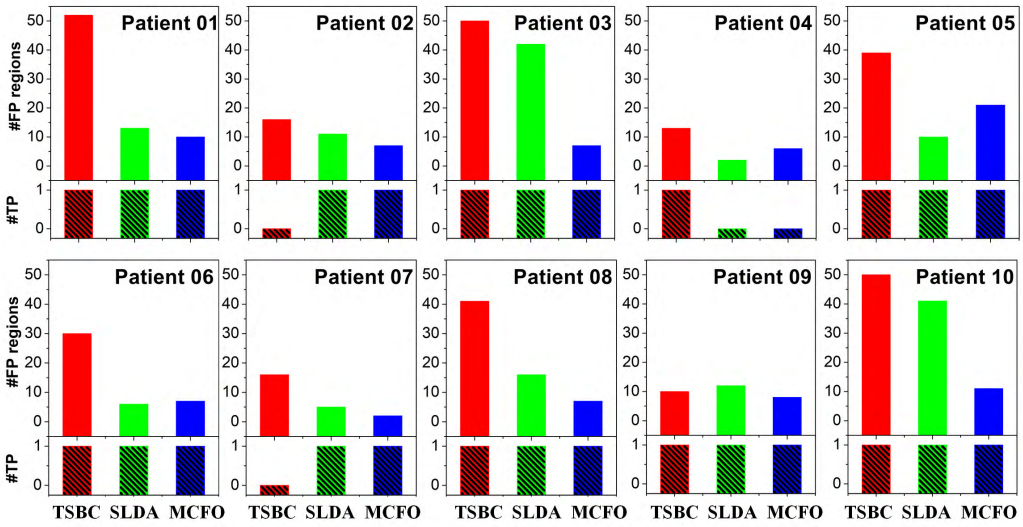

FIGURE 7. Graphical representation of Table 4: evaluation results of FCD detection algorithms that use region-based analysis. For each patient, the top plot shows the number of FP regions detected, and the bottom plot shows whether the lesion (TP region) was detected.

four patients and vividly illustrate the benefits of the lower false-positive results.

The voxel-based evaluation (Table 3) revealed various interesting trends. First, all three methods had an accuracy of 0.94 or higher, although this could be a misleading metric because the number of TN voxels was significantly higher than the number of the TP, FP, and FN voxels. Therefore, the precision, recall, and F-score were more informative. TSBC underperformed based on all three metrics. MCFO achieved a better F-score and higher precision in comparison with SLDA, although it achieved a lower TPR than SLDA. However, because MCFO was able to detect the same lesions as those detected by SLDA at the regional level (Table 4), the lower TPR, in terms of detected voxels, suggests that MCFO is more conservative in identifying the lesion boundaries. This phenomenon can also be observed in Figs. 8 and 9. The lesion boundaries identified by the MCFO were slightly smaller than those identified by the SLDA and the ground truth.

The benefit of MCFO can be seen in the high precision and low FPR values presented in Table 3, and the low number of FP regions presented in Table 4 and Fig. 7. The number of FP regions per patient from MCFO, SLDA and TSBC was 8.6, 15.8, and 31.7, respectively. MCFO produced the fewest FP results both in terms of voxels and in terms of regions, which is beneficial to clinicians who need to evaluate each candidate FCD region. In other words, the proposed scheme is more likely to recognize non-lesional regions as genuinely normal brain tissue, in comparison with the other two methods. Specifically, the precision of MCFO was approximately 82% and 43% higher, on average, in comparison with TSBC and SLDA, respectively, whereas its FPR was 88% and 70% lower than the others, respectively. In 7 out of 10 patient images, MCFO produced the fewest FP regions (between 2 and 11 in total). In the other three patient images, MCFO identified between 1 and 11 more FP regions than SLDA, which is still much less than TSBC.

The qualitative comparison of two methods for FCD detection (TSBC and SLDA) and the proposed MCFO method are

TABLE 4. Evaluation of FCD detection algorithms using region-based analysis.

	Lesion Detected			#FP regions (R_{min}=900=112.5mm^3)		
Patient	TSBC	SLDA	MCFO	TSBC	SLDA	MCFO
1	Y	Y	Y	52	13	10
2	**N**	Y	Y	16	11	7
3	Y	Y	Y	50	42	7
4	Y	**N**	**N**	13	2	6
5	Y	Y	Y	39	10	21
6	Y	Y	Y	30	6	7
7	**N**	Y	Y	16	5	2
8	Y	Y	Y	41	16	7
9	Y	Y	Y	10	12	8
10	Y	Y	Y	50	41	11

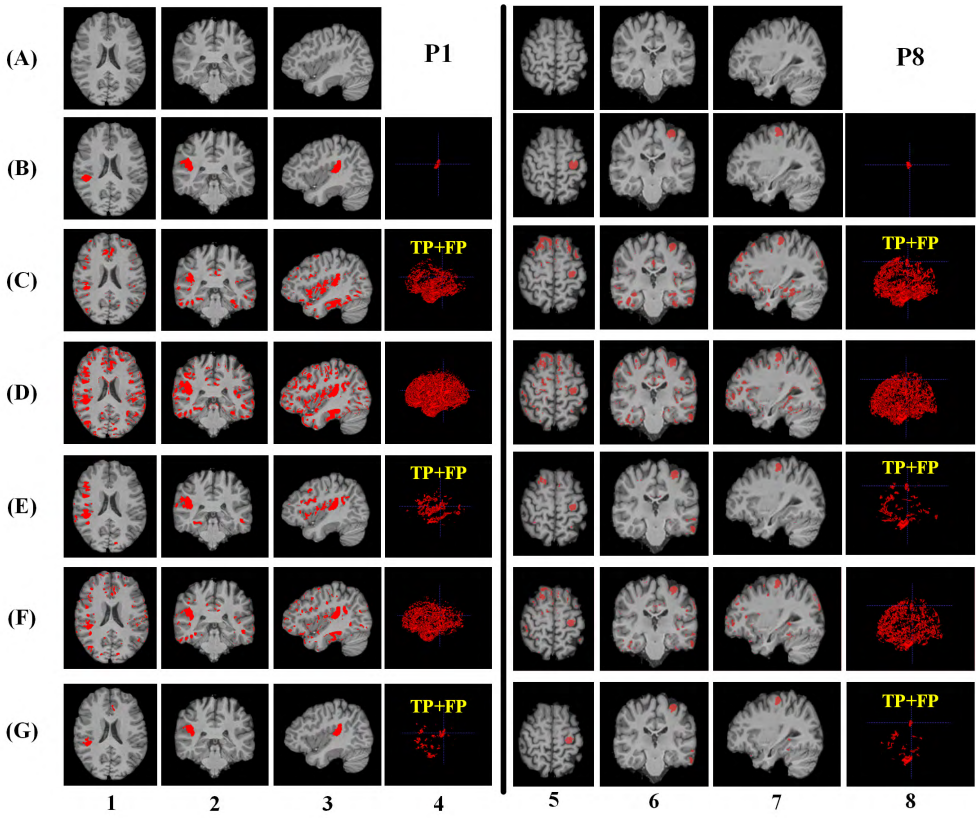

FIGURE 8. Images are the classification results of patients 1 (P1) and 8 (P8). Rows are the preprocessed images (A), ground truth with lesional region indicated in red (B), classification results by TSBC (C), voxel-based classification results of SLDA (D), region-based classification of SLDA (E), voxel-based classification results of MCFO (F), and region-based classification of MCFO (G). Columns 1-3 and 5-7 represent the axial, sagittal, and coronal views of the images; regions in red indicate possible lesions. Columns 4 and 8 represent the volumetric views of the regions identified as positive voxels.

presented in Fig. 8. For a qualitative comparison, the two existing methods of FCD detection (TSBC and SLDA) were reproduced to process the datasets used in this study. For brevity, the patient numbers 1 (P1) and 8 (P8) are plotted in Fig. 8 to illustrate the different levels of positive voxel detection. The ground truth [row (B)] is referenced to compare the final classified results of TSBC (row C), SLDA (row E) and MCFO (row G).

In the two-dimensional illustration (columns 1-3, and 5-7) shown in Fig. 8, close matches were observed between the lesion regions that were correctly identified as positive voxels by all methods (rows C, E, G) and the ground truth (row B). Additionally, we observed that the regional healthy cortices were misinterpreted as positive voxels, and were thus regarded as false-positive results. For example, as shown in the lower right corner of Fig. 8 $(C, E, G)_6$, all methods falsely identified the healthy region near the bottom of the brain tissue as positive.

For the 3D illustrations, the results of the proposed method shown in Fig. $8(G)_{4,8}$ reproduced an approximately identical shape of the lesion as the ground truth (Fig. $8(B)_{4,8}$), where the false-positive regions were much less than those for TSBC (Fig. $8(C)_{4,8}$) and SLDA (Fig. $8(E)_{4,8}$). This finding is supported by the following observations: 1) for P1, the voxel-based FPR of the TSBC, SLDA, and MCFO was 0.0965, 0.0343 and 0.0072, respectively; 2) for P8, the voxel-based FPR by TSBC, SLDA, and MCFO was 0.0616, 0.0203 and 0.0088, respectively. Even if the lesion in P8 was half the size of P1, the proposed MCFO also produced fewer false-positive results over TSBC and SLDA, which evidently validates the capability of the MCFO scheme in terms of false-positive minimization. The 3D illustrations also reveal that the final results of all methods included false positives and true positives (yellow TP+FP in columns 4 and 8).

The regional classification of SLDA (Fig.$8(E)_{4,8}$) and MCFO (Fig.$8(G)_{4,8}$) both effectively reduced the false-positive results obtained by the voxel-based classification (Fig.$8(D, F)_{4,8}$). Therefore, the regional classification results were considered as the final results of SLDA and MCFO (Fig. 9). After re-labeling the noise and tiny fragmental region (which was smaller than R_{min}) as negative, no region was classified as lesional by the SLDA and MCFO in the images of 31 healthy controls. For the TSBC method, neither the standard pre-surgical evaluation nor the classifier were successful in identifying lesions in healthy control subjects [3].

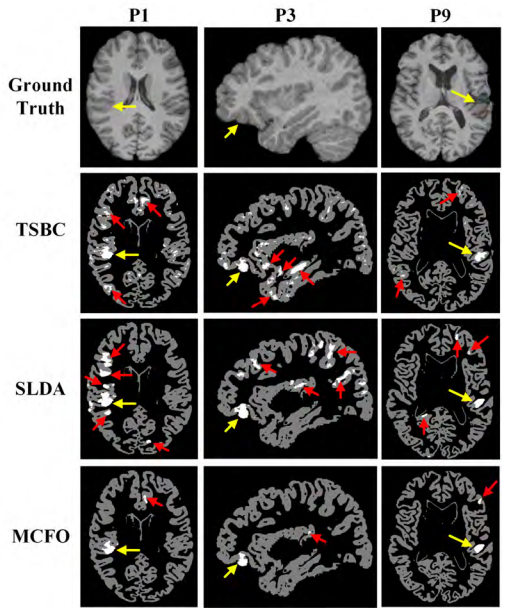

FIGURE 9. Example region-based output from the three FCD detection algorithms superimposed upon an example 2D slice from three patients. The first row is the ground truth image with the FCD lesion indicated by the yellow arrow. Rows 2-4 are the classified images from SLDA, TSBC and the proposed MCFO, respectively, as superimposed on the GM segmentation. Yellow and red arrows indicate the true- and false-positive regions found by the algorithms, respectively. These examples show the benefit of MCFO that fewer FP regions are generated.

E. RESULTS OF EXPERTS' EVALUATIONS

When only T1 images were used, one radiologist named J. Guo (with 14 years of work experience) correctly identified 2/10 patients (true positive) and incorrectly identified 1/10 healthy controls (false positive) as positive. Another radiologist Q. Wang (with 4 years of work experience) correctly identified 3/10 patients (true positive) and incorrectly identified 4/10 healthy controls as positive (false positive). By combining with T1 and FLAIR, J. Guo found lesions for 8/10 patients (true positives) and for 4/10 healthy controls (false positives), while Q. Wang found lesions for 9/10 patients (true positives) and for 6/10 healthy controls (false positive). For Patient 04, both radiologists failed to find the lesional location in T1 images, because the lesion was very small and there were no visual features on the images.

The following conclusions are drawn from these results: the FCD structure features in the T1 images are difficult to be identify with a naked eye; the FCD features in FLAIR lack specificity, and healthy controls also have similar hyperintensity; automated method could help improve the detection of FCD structural changes in T1.

IV. DISCUSSION

Our study demonstrates that the proposed MCFO framework is a promising tool for automatically detecting FCD lesions, because it achieves the same or better detection performance as two state-of-the-art FCD detectors (SLDA and TSBC), while generating fewer FP candidate lesional regions. The proposed framework could identify more patients with pharmacoresistant epilepsy suffering from FCD, who could benefit from surgical treatment. Although the diagnosis of FCD, and consequently the decision to conduct epilepsy surgery, can never rely on a single diagnostic tool [12], the MCFO method could serve as an additional screening tool by providing the candidate positive regions in the location of possible dysplasia. Considering that the diagnosing doctor must evaluate each candidate positive region, the generation of fewer false positive results can save time and increase the doctors' diagnostic efficiency.

In this study, we propose the use of the F-score to construct the objective function, which considers only the number of TP, FN, and FP voxels, in the objective function. This concept is contrary to the *accuracy* metric typically used in classification, which includes the number of TN voxels. In our FCD dataset, the number of TN voxels is on average 30 times greater than all other voxel classes, which renders accuracy useless as an optimization function. Likewise, when evaluating the detection results, the accuracy metric is not sufficiently informative. The use of the F-score as the objective function also minimizes the number of FPs, and such a minimization is an essential step toward the objective of reducing the doctors' workload.

The proposed method is based on an optimized multiple classifier model, in contrast to previous attempts that have relied on single classifiers. The selected approach combines the strengths of different classifiers in one framework [21]. The results of the comparison between the proposed MCFO and existing methods (SLDA and TSBC) revealed that the combination of multiple classifiers improves the performance of FCD detection in comparison with the performance of a single classifier.

The classification scheme proposed in this study relies both on feature maps and their corresponding Z-score maps. Although the feature maps measure the difference between the lesional and non-lesional regions within the images of one person (intra-subject), the Z-score maps measure the variation in comparison with healthy controls (inter-subject). Originally, we hypothesized that the Z-score maps would result in better lesion detection in comparison the feature maps. However, not all features had higher performing Z-score images:

1) Human brain structures are complex and aligning the brain structures of different individuals is difficult.

2) Comparing the brains of different subjects is challenging. The FCD lesions are related to abnormal structural changes and local intensity changes. If a region in the feature map is markedly different from the same region for healthy controls, then, the probability of classifying the region as positive is higher. However, the brain structures of different subjects are inherently or naturally different.

By sacrificing small parts of the lesional region, the proposed MCFO achieved a greater reduction of FPR in comparison with TSBC and SLDA (Table 3). The TPR was equal to the percentage overlap (coverage) between the lesional mask

in ground truth and the detected region. The MCFO detected more FCD than TSBC (0.4500 for MCFO and 0.3933 for TSBC), and slightly less than SLDA (0.5374 for SLDA), as presented in Table 3. However, the FPR was much less than that of SLDA (0.0064 for MCFO and 0.0212 for SLDA). With regarded to lesional detection, there is a trade-off between the coverage rate and false positive.

The proposed MCFO achieved an improvement in most of the subjects, but not all subjects (Fig.7). For 7/10 patients (Patient 01, 02, 03, 07, 08, 09, and 10), the proposed MCFO had a lower number of FP regions than TSBC and SLDA. For the 3/10 patients (Patient 04, 05, and 06), the proposed MCFO had a higher number of FP regions in comparison with SLDA. The high FP rate occurred because of the following reasons: 1) the FCD lesional region can be very tiny (500 mm^3 for Patient 04), and the features of these miniscule lesional regions are similar to those of non-lesional regions. For example, the cortical thickness value in the miniscule lesional region may be 5 mm, and this value could also appear in a non-lesional region for all subjects. Therefore, to capture the features of these miniscule lesional regions, the classifiers inevitably misclassified the non-lesional region as lesional. Thus, more effective features must be identified in future work. 2) The variation of lesional size is large (ranges from 500 mm^3 to 23670 mm^3), which leads to the training datasets not being able to effectively model the features of the test datasets. By increasing the number of datasets and analyses, the FCD in different size levels may further reduce the number of FP regions. 3) The registration accuracy also affects the FP results. To model the changes of patients relative to the healthy controls, the computation of the z-score of feature images crosses different subjects, and the inter-subject variability of the human brain affects the comparison of the obtained results.

To mirror the clinical context, we conducted a regional assessment of the classification results, in addition to performing analysis at the voxel level. By contrast, the overall volumes of the identified lesions were, on average, smaller with MCFO, as suggested by the lower number of TP voxels identified per lesion. This observation suggests that our method is as effective as SLDA, and better than TSBC, at identifying the true lesional regions. Considering that our ground truth was determined by a single neuroradiologist, any manual segmentation errors could have affected the voxel-based operations (training and evaluation). However, the primary objective of FCD algorithms is to improve detection and visualization of FCD lesions, and not to perform surgical planning (electroencephalography is preferred for this purpose). Therefore, the imperfect delineation of the FCD boundary in MRI by the automated methods may be acceptable. The final clinical decisions will be taken by a multidisciplinary team, including a neuroradiologist, and will not be solely based on an algorithm.

Moreover, it may be unfair for a radiologist to examine the images without information regarding clinical symptoms, history of diseases, and demographic data (age and sex), because radiologists typically diagnose diseases by using images and other information related to the suspected disease. In the future, if machine learning classifiers make decisions using MRI images along with other important information such as clinical symptoms, this method will become more practical.

This study relied only on T1-weighted images for FCD detection, and did not consider FLAIR. The primary reason for this is that the gray matter and white matter had good contrast on the T1-weighted images, and this rendered them more suitable for capturing the structural changes in the lesional region of the gray and white matter. However, multimodal image analysis has considerable potential for improving the detection performance of the FCD lesions. Gill *et al.* [36] enhanced the co-occurring FLAIR hyperintensity and the T1 hypointensity occurring at the junction between the gray and white matter, and correctly identified 34/41 lesions and 35/38 controls. Adler *et al.* [17] combined the features of the T1 images and FLAIR images, and achieved a sensitivity of 73% for FCD detection in a pediatric cohort. The lesional features may do not appear in the same location in T1 and FLAIR. For example, the increased cortical thickness feature of T1 is within the cortex, whereas the FLAIR hyperintensity could extend from the cortex to the brain region near the ventricle. Surface-based detection could solve the problem of non-overlapping from multimodal features, because it maps all features on the cortical surface. However, surface-based detection cannot fully delineate the lesions, but voxel-based detection can. Generally, surface-based detection has good potential for FCD detection using a multimodal MRI.

The dataset used in this study was limited in terms of data quantity. Therefore, leave-one-out cross validation was used to evaluate the results (each image was used as a test image, and training was carried out using the rest of the images). The low prevalence of FCD hinders the development of a sufficiently large local testing dataset. Therefore, we attempted to find publicly available MRI datasets containing FCD lesions. To our knowledge, only one dataset of epilepsy patients[4] is available for public research, and only three cases in that dataset contain FCD lesions. However, these patients are infants, whose brain tissue is immature; therefore, these brain images are unsuitable for use within the scope of this study. Kini *et al.* [7] have cited the lack of public FCD datasets as a significant problem in FCD research.

A limited dataset cannot represent all varieties of type II FCD, and affects the classification parameters and the validation methods. For example, the naïve Bayesian classifier uses the feature value distributions, when the number of datasets increases, and the feature distribution may tend to provide a better representation of type II FCD. A large number of datasets can be randomly separated into training and testing data, while a limited dataset cannot be separated and must be analyzed using leave-one-out validation. Additionally, if the

[4] http://eeg.pl/epi

number of dataset is large, a deep-learning method is also a promising option.

V. CONCLUSIONS

In this study, an MCFO scheme is proposed to reduce the number of false positives during the detection of FCD lesions. The proposed method fuses the outputs from multiple classifiers to rank and select candidate lesional features. These features are identified by using a constructed object function, which maximizes the F-score of lesion detection. A genetic algorithm yields the optimal weightings of multiple classifiers, and achieves a marked reduction in the number of detected false positives. The proposed scheme achieved a 9/10 subject-wise lesion recognition rate, with a much lower quantity of false-positive regions in comparison with those of TSBC and SLDA.

In this study, the multiple classifier strategy explored a finer framework for the automation of neuropathological diagnosis. The reduced number of false positive candidate lesions, along with the high true-positive recognition rate, can reduce the time needed for doctors to identify a true lesion amongst the positive results provided by an automated detection method.

LIST OF ACRONYMS

FCD	focal cortical dysplasia
MR	magnetic resonance
FP	false-positive
TP	true positives
TN	true negative
FN	false-negative
SD	standard deviation
MCFO	multiple classifier fusion and optimization
GM/WM	gray matter/white matter
CSF	cerebral spinal fluid
FLAIR	fluid attenuated inversion recovery
VBM	voxel-based morphometry
NB	naïve Bayesian
NN	neural network
SVM	support vector machine
TSBC	two-stage Bayesian classifier
SLDA	surface features combined with a linear discriminant analysis
MCS	Multi-classifier systems
FAST	FMRIB's Automated Segmentation Tool
MDA	mahalanobis discriminant analysis
QDA	quadratic discriminant analysis

REFERENCES

[1] A. Bernasconi, N. Bernasconi, B. C. Bernhardt, and D. Schrader, "Advances in MRI for 'cryptogenic' epilepsies," *Nature Rev. Neurol.*, vol. 7, pp. 99–108, Jan. 2011.

[2] I. Despotović *et al.*, "Automatic 3D graph cuts for brain cortex segmentation in patients with focal cortical dysplasia," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Boston, MA, USA, Aug./Sep. 2011, pp. 7981–7984.

[3] S. B. Antel *et al.*, "Automated detection of focal cortical dysplasia lesions using computational models of their MRI characteristics and texture analysis," *NeuroImage*, vol. 19, pp. 1748–1759, Aug. 2003.

[4] I. Blümcke *et al.*, "The clinicopathologic spectrum of focal cortical dysplasias: A consensus classification proposed by an ad hoc task force of the ILAE diagnostic methods commission," *Epilepsia*, vol. 52, pp. 158–174, Jan. 2011.

[5] K. Deblaere and E. Achten, "Structural magnetic resonance imaging in epilepsy," *Eur. Radiol.*, vol. 18, pp. 119–129, Jan. 2008.

[6] P. Krsek *et al.*, "Different features of histopathological subtypes of pediatric focal cortical dysplasia," *Ann. Neurol.*, vol. 63, pp. 758–769, Jun. 2008.

[7] L. G. Kini, J. C. Gee, and B. Litt, "Computational analysis in epilepsy neuroimaging: A survey of features and methods," *NeuroImage, Clin.*, vol. 11, pp. 515–529, Feb. 2016.

[8] A. Bernasconi *et al.*, "Texture analysis and morphological processing of magnetic resonance imaging assist detection of focal cortical dysplasia in extra-temporal partial epilepsy," *Ann. Neurol.*, vol. 49, pp. 770–775, Jun. 2001.

[9] H.-J. Huppertz *et al.*, "Enhanced visualization of blurred gray–white matter junctions in focal cortical dysplasia by voxel-based 3D MRI analysis," *Epilepsy Res.*, vol. 67, pp. 35–50, Oct./Nov. 2005.

[10] P. Besson, F. Andermann, F. Dubeau, and A. Bernasconi, "Small focal cortical dysplasia lesions are located at the bottom of a deep sulcus," *Brain*, vol. 131, pp. 3246–3255, Dec. 2008.

[11] L. Bonilha *et al.*, "Voxel-based morphometry reveals excess gray matter concentration in patients with focal cortical dysplasia," *Epilepsia*, vol. 47, pp. 908–915, May 2006.

[12] J. Kassubek, H.-J. Huppertz, J. Spreer, and A. Schulze-Bonhage, "Detection and localization of focal cortical dysplasia by Voxel-based 3-D MRI analysis," *Epilepsia*, vol. 43, pp. 596–602, Jun. 2002.

[13] P. Besson, O. Colliot, A. Evans, and A. Bernasconi, "Automatic detection of subtle focal cortical dysplasia using surface-based features on MRI," in *Proc. 5th IEEE Int. Symp. Biomed. Imag., Nano Macro (ISBI)*, Paris, France, May 2008, pp. 1633–1636.

[14] C. Loyek, F. G. Woermann, and T. W. Nattkemper, "Detection of focal cortical dysplasia lesions in MRI using textural features," in *Proc. Bildverarbeitung für die Medizin*, 2008, pp. 432–436.

[15] S.-J. Hong, H. Kim, D. Schrader, N. Bernasconi, B. C. Bernhardt, and A. Bernasconi, "Automated detection of cortical dysplasia type II in MRI-negative epilepsy," *Neurology*, vol. 83, pp. 48–55, Jun. 2014.

[16] T. Thesen *et al.*, "Detection of epileptogenic cortical malformations with surface-based MRI morphometry," *PLoS ONE*, vol. 6, no. 2, p. e16430, 2011.

[17] S. Adler *et al.*, "Novel surface features for automated detection of focal cortical dysplasias in paediatric epilepsy," *NeuroImage, Clin.*, vol. 14, pp. 18–27, Jan. 2017.

[18] T. M. Schwartz *et al.*, "The interplay of attention economics and computer-aided detection marks in screening mammography," *Proc. SPIE*, vol. 9787, p. 978711, Mar. 2016.

[19] W. Jorritsma, F. Cnossen, and P. M. A. van Ooijen, "Improving the radiologist–CAD interaction: Designing for appropriate trust," *Clin. Radiol.*, vol. 70, pp. 115–122, Feb. 2015.

[20] C. K. Chow, "Statistical independence and threshold functions," *IEEE Trans. Electron. Comput.*, vol. EC-14, no. 1, pp. 66–68, Feb. 1965.

[21] M. Wozniak, M. Grana, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," *Inf. Fusion*, vol. 16, pp. 3–17, Mar. 2014.

[22] C. R. Houck, J. A. Joines, and M. G. Kay, "A genetic algorithm for function optimization: A MATLAB implementation," North Carolina State Unvi., Raleigh, NC, USA, Tech. Rep. NCSUIE-TR-95, 2008.

[23] X. Qu *et al.*, "Local directional probability optimization for quantification of blurred gray/white matter junction in magnetic resonance image," *Frontiers Comput. Neurosci.*, vol. 11, p. 83, Sep. 2017.

[24] X. Qu, J. Yang, S. Ma, T. Bai, and W. Philips, "Positive unanimous voting algorithm for focal cortical dysplasia detection on magnetic resonance image," *Frontiers Comput. Neurosci.*, vol. 10, p. 25, Mar. 2016.

[25] X. Qu *et al.*, "Estimating blur at the brain gray-white matter boundary for FCD in MRI," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBS)*, Chicago, IL, USA, Aug. 2014, pp. 3321–3324.

[26] P. A. Yushkevich *et al.*, "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability," *NeuroImage*, vol. 31, no. 3, pp. 1116–1128, Jul. 2006.

[27] S. M. Smith, "Fast robust automated brain extraction," *Hum. Brain Mapping*, vol. 17, pp. 143–155, Nov. 2002.

[28] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm," *IEEE Trans Med. Imag.*, vol. 20, no. 1, pp. 45–57, Jan. 2001.
[29] M. Jenkinson, P. Bannister, M. Brady, and S. Smith, "Improved optimization for the robust and accurate linear registration and motion correction of brain images," *NeuroImage*, vol. 17, no. 2, pp. 825–841, Oct. 2002.
[30] M. Jenkinson and S. Smith, "A global optimisation method for robust affine registration of brain images," *Med. Image Anal.*, vol. 5, no. 2, pp. 143–156, Jun. 2001.
[31] O. Lyttelton, M. Boucher, S. Robbins, and A. Evans, "An unbiased iterative group registration template for cortical surface analysis," *NeuroImage*, vol. 34, pp. 1535–1544, Feb. 2007.
[32] P. Besson, N. Bernasconi, O. Colliot, A. Evans, and A. Bernasconi, "Surface-based texture and morphological analysis detects subtle cortical dysplasia," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI).*, vol. 11, 2008, pp. 645–652.
[33] Y. Chen *et al.*, "Bayesian statistical reconstruction for low-dose X-ray computed tomography using an adaptive-weighting nonlocal prior," *Comput. Med. Imag. Graph.*, vol. 33, no. 7, pp. 495–500, 2009.
[34] Y. Chen, J. Ma, Q. Feng, L. Luo, P. Shi, and W. Chen, "Nonlocal prior Bayesian tomographic reconstruction," *J. Math. Imag. Vis.*, vol. 30, no. 2, pp. 133–146, 2008.
[35] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. Hoboken, NJ, USA: Wiley, 2000.
[36] R. S. Gill *et al.*, "Automated detection of epileptogenic cortical malformations using multimodal MRI," in *Proc. Int. Workshop Deep Learn. Med. Image Anal., Int. Workshop Multimodal Learn. Clin. Decis. Support*, 2017, pp. 349–356.

XIAOXIA QU received the Ph.D. degree in optical engineering from the School of Optics and Photonics, Beijing Institute of Technology, Beijing, China, in 2017. She has been a Joint-Ph.D. Student with Ghent University, Ghent, Belgium, for two years, supported by the China Scholarship Council. She has been a Researcher and an Engineer with the Radiology Department, Beijing Tongren Hospital, Capital Medical University, Beijing, since 2017. Her research interests include brain magnetic resonance imaging segmentation and pattern recognition.

JIAN YANG received the Ph.D. degree in optical engineering from the School of Optics and Photonics, Beijing Institute of Technology, Beijing, China, in 2007. He held a post-doctoral position with Children's Hospital Toronto, Canada, from 2007 to 2009. He has been with the Beijing Institute of Technology since 2009, where he is currently a Full Professor with the School of Optoelectronics. His research interests include medical image processing, computer vision, and artificial intelligence.

LJILJANA PLATIŠA received the M.S. degree from the Faculty of Technical Sciences, University of Novi Sad, Serbia, in 2001, and the Ph.D. degree in computer science engineering from Ghent University, Ghent, Belgium, in 2014. From 2004 to 2007, she was a Project Manager and the Chief Executive Officer with Zesium mobile d.o.o. She is currently a Post-Doctoral Researcher with imec-IPI-UGent. Her research interests include image quality evaluation and digital image analysis.

ASLI KUMCU received the M.S. degree in medical imaging from the University of Leuven in 2008. She was with Barco's Healthcare Division for five years. She has been a Researcher with Ghent University since 2010. Her research interests are subjective and objective image/video quality assessment methods and their use in optimizing and validating the image quality of medical imaging devices.

DANNI AI received the B.S. and M.S. degrees from Xi'an Jiaotong University in 2008 and 2005, respectively, and the Ph.D. degree from Ritsumeikan University, Japan, in 2011. She is currently an Associate Professor with the School of Optics and Photonics, Beijing Institute of Technology. Her research interests include medical image analysis, surgical navigation, virtual reality, and augmented reality.

BART GOOSSENS received the M.S. degree in computer science and the Ph.D. degree in engineering from Ghent University in 2006 and 2010, respectively. He is currently a Professor of digital image processing with the Image Processing and Interpretation Research Group, Department of Telecommunications and Information Processing. His research interests include medical image reconstruction (CT and magnetic resonance image), noise modeling and estimation, and medical image quality assessment.

TINGZHU BAI received the Ph.D. degree in optical engineering from the School of Optics and Photonics, Beijing Institute of Technology, Beijing, China, in 2001. He is currently a Professor with the School of Optoelectronics, Beijing Institute of Technology. His research interests include image segmentation and pattern recognition.

YONGTIAN WANG received the Ph.D. degree from the Department of Physics, University of Reading, U.K., in 1986. He has been with the Beijing Institute of Technology since 1988, where he is currently a Professor with the School of Computer Science and the School of Optoelectronics. His research interests include computer science, computer vision, medical image processing, and image guided surgery.

JING SUI received the Ph.D. degree in optical engineering from the School of Optics and Photonics, Beijing Institute of Technology, Beijing, China, in 2007. She is currently a Professor with the Institute of Automation, Chinese Academy of Sciences. Her research interests are data mining, machine learning, pattern recognition, image processing/denoising, and high-dimensional data analysis.

KAREL DEBLAERE received the Ph.D. degree from Ghent University, Belgium, in 2004. He is currently a Professor with the Department of Radiology and Nuclear Medicine, Ghent University. He is also a Neuroradiologist with Ghent University Hospital. His research interests include brain magnetic resonance image (MRI) analysis, PET for therapy response assessment, medical physics, and fMRI.

WILFRIED PHILIPS was born in Aalst, Belgium, in 1966. He received the Diploma and Ph.D. degrees in electrical engineering from the University of Ghent, Belgium, in 1989 and 1993, respectively. From Oct. 1989 to Sep. 1993, he was working as a research assistant at the Department of Electronics and Information Systems of Ghent University, for the Flemish Fund for Scientific Research (FWO). He is currently a Senior Professor and the Leader with the Image Processing and Interpretation Research Group, TELIN Department, Ghent University. His main research interests are image and video restoration, image analysis, and lossless and lossy data compression of images and video and processing of multimedia data.